1995

Priorities of Quality

Carol Tenopir
University of Tennessee - Knoxville

Recommended Citation

Carol Tenopir. "Priorities of Quality," in: Electronic Information Delivery: Ensuring Quality and Value, Reva Basch, ed. London: Gower, 1995. Pp. 119-139.

Chapter 8

PRIORITIES OF QUALITY

Carol Tenopir, Professor at the University of Hawaii at Manoa School of Library and Information Studies

Introduction

Many groups and individuals are looking at ways databases can be evaluated for quality in a consistent way. Several have developed detailed checklists of database quality to help with this evaluation process.[1] Three of the best are those developed by:

- the Southern California Online User's Group (SCOUG)[2]
- the Finnish Society for Information Services[3]
- Peter Jacsó, in his 1992 book on CD-ROM evaluation.[4]

SCOUG's proposed rating scheme for bibliographic, full-text and directory databases focuses on evaluating the results or outputs from an online search. They identified 11 main components by which a searcher can rate quality: consistency; coverage/scope; timeliness; value in terms of cost; accuracy/error rate; accessibility; system performance/ease of use; integration with other databases; output; documentation; and customer support.[5]

Because these factors will vary with the system that loads a database, SCOUG recommended that every system should be evaluated separately. Thus, for a database such as ABI/INFORM that is available on multiple online or CD-ROM hosts, multiple SCOUG evaluations and ratings are required.

The Finnish Society for Information Services' rating scheme focuses just on bibliographic databases, proposing five main categories of evaluation. The five categories to evaluate are: connecting to system and communications; search

language and other technical aspects of the search; contents quality; aids to information retrieval; and costs.[6]

'Contents quality' in the Finnish scheme includes over 30 separate points. These can be roughly grouped into points covering: update policies; coverage policies; indexing and abstracting and other value-added fields; consistency of word forms; errors; and standardization of abbreviations.

Jacsó's book, although focusing on CD-ROM databases, contains evaluation criteria applicable to any electronic medium. His criteria are separated into four broad categories: software, dataware, database and hardware.[7] Under dataware he includes scope, content and quality; under database he includes accessibility, documentation/user support, installation and terms (including costs and licensing).

Each of these three proposes a checklist approach that will lead to consistent and systematic evaluation. SCOUG and Jacsó recommend computing a numeric 'quality' score, so similar databases can be compared and database producers and system vendors will see how their offerings rate.

Such checklists are valuable because they help codify, clarify and compartmentalize various aspects of quality. They ensure that consistent criteria are used when evaluating and comparing different databases, no matter the electronic medium.

What such checklists do not usually do is prioritize on a general level, although Jacsó's evaluation forms let each consumer set a weight for each factor to be evaluated. Most checklists make all quality issues appear equal, from a poor telephone connection, to a typographical error in the abstract, to a date field missing in many records in a database, to inadequate search software.

The other thing checklists often do is focus on *comparison* rather than on *quality*. For example, the size of a database is listed as one component of quality in some lists. *Size* is a comparative factor, not a quality one. One database may be described as having more records than another, but that says nothing about the quality of either. Bigger is not necessarily better. Comparison factors are relative to another database, quality is not relative.

This chapter builds on the recent checklists with a focus that is a bit deeper and narrower. It will examine what is *most* important in both defining and making a quality database and will concentrate on those things that are absolute quality factors, whether an evaluator is comparing databases or just looking at a single database.

Aspects of quality

As these various checklists illustrate, there are many different aspects that contribute to a quality experience in a database search. Martha E. Williams presents the aspects as an information generation–database use chain that focuses on the people responsible for quality at each link. Her seven-link chain includes: author/originator, primary publisher, secondary publisher (database producer), tertiary publisher (online vendor), gateway, searcher/analyst, and enduser/requestor.[8]

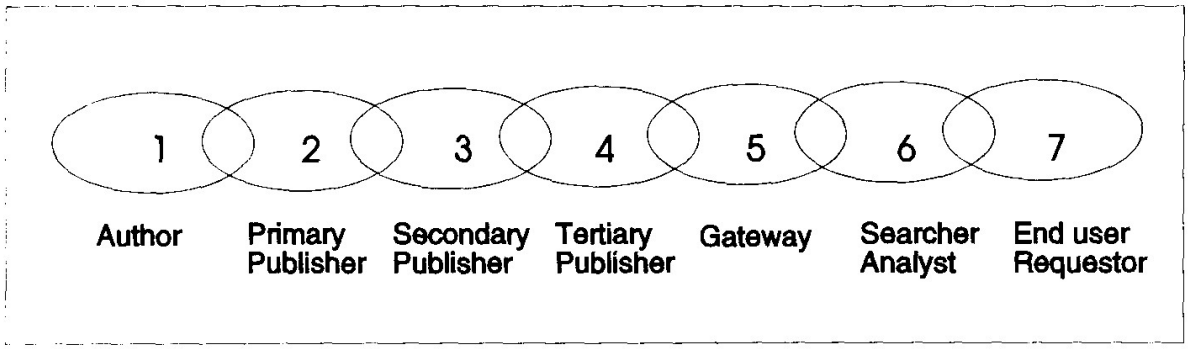

Figure 8.1 *Information generation – database use chain*

A more common way to look at quality is by component parts. SCOUG saw just three parts: the data itself, the database structure and organization, and the system on which it is loaded.[9] Jacsó calls these three dataware, database, and software.[10]

Figure 8.2 diagrams the components of a database system and where they can be separated.[11] The information content is the information that is gathered or created by the publisher or producer. Secondary publishers of indexes and abstracts arrange

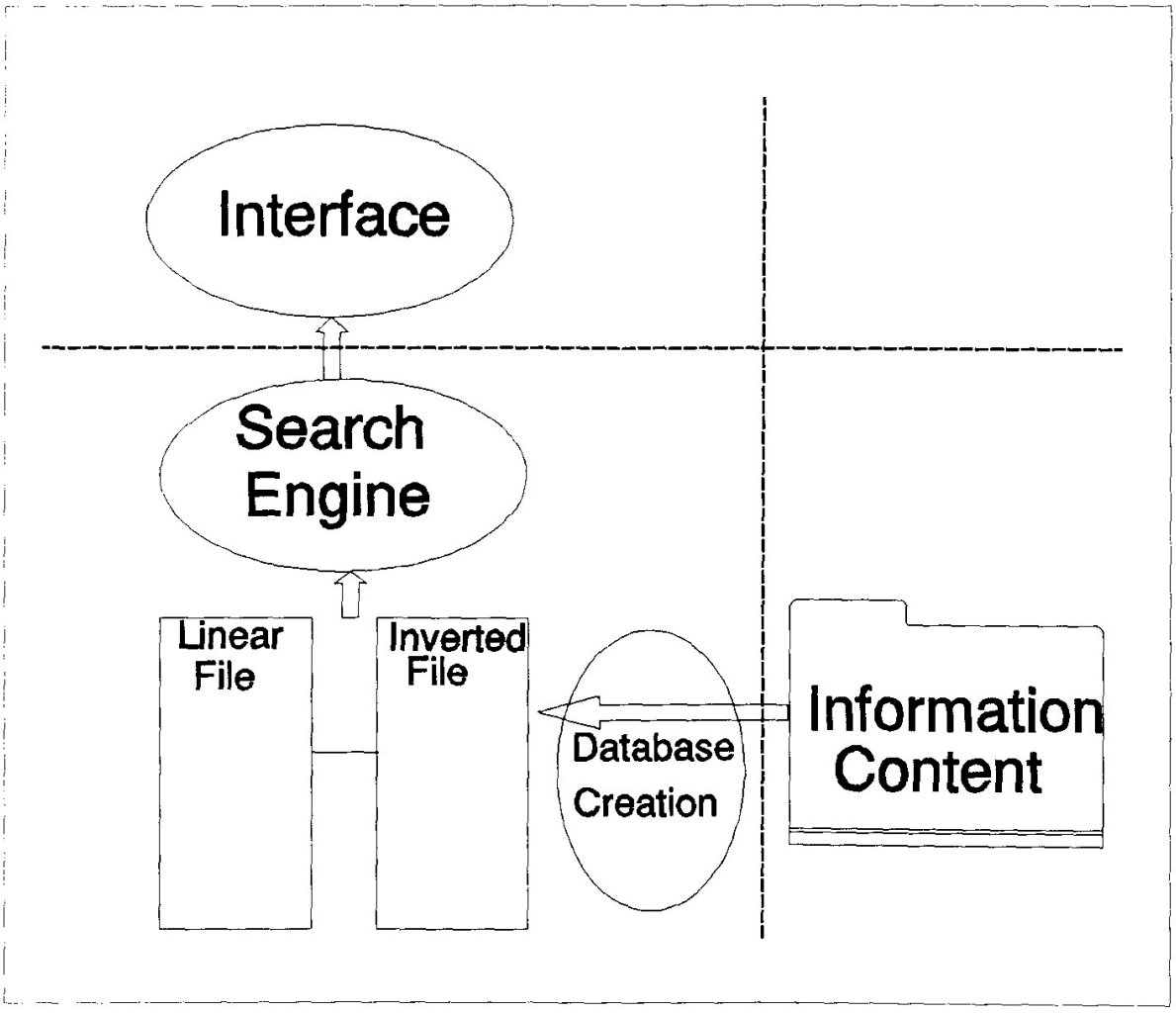

Figure 8.2 *Separation of the components of a database system*

bibliographic information and add value, such as abstracts or descriptors, to it. The information content is independent of software and of the distribution medium (print, CD-ROM, online or other). Content may be evaluated separately, as the dotted line indicates.

System software transforms the information content into a database by making the information searchable, creating inverted indexes (dictionary files) for searching with pointers to the linear file that is used to display records.

The retrieval software, also called the search engine, contributes search features that allow the database indexes to be searched. The power of retrieval software varies from system to system, a major reason that SCOUG recommends evaluating each system's version of a database separately.

The final component of a database is the user/system interface. It is how the user interacts with the retrieval software to use the database. The interface may use commands, menus, icons, templates, function keys, or a combination of these. Because several online and CD-ROM systems offer a choice of interfaces, the interface can be evaluated separately from the other database aspects.

Components or aspects of quality can also be categorized by function, as the Finnish Society for Information Services did with its five main functions.[12] Functions can also be viewed as logging on, searching, printing, and evaluating or using results.

Core of quality

The quality database searching experience depends on reliable telecommunications links, good, powerful (and easy) search software, predictable and acceptable response time, and a competent searcher, but all of these may be present without the total adding up to quality. One factor is central, without which quality cannot exist. At the core of all aspects is content – the information itself as created by the author, secondary publisher or database producer. Without quality content, the other aspects become unimportant. In that sense, content is the highest priority aspect.

Responsibility

Who is responsible for database content quality? Williams's chain implies that all seven links are equally responsible, but a perusal of the online system terms and conditions-of-use that include database producer disclaimers suggests that *no-one* is responsible!

Here are excerpts from a few typical disclaimers in Dialog's terms and conditions:[13]

> BIOSIS does not guarantee the accuracy, adequacy, or completeness of any information. ... BIOSIS disclaims all liability for errors or omissions ... .

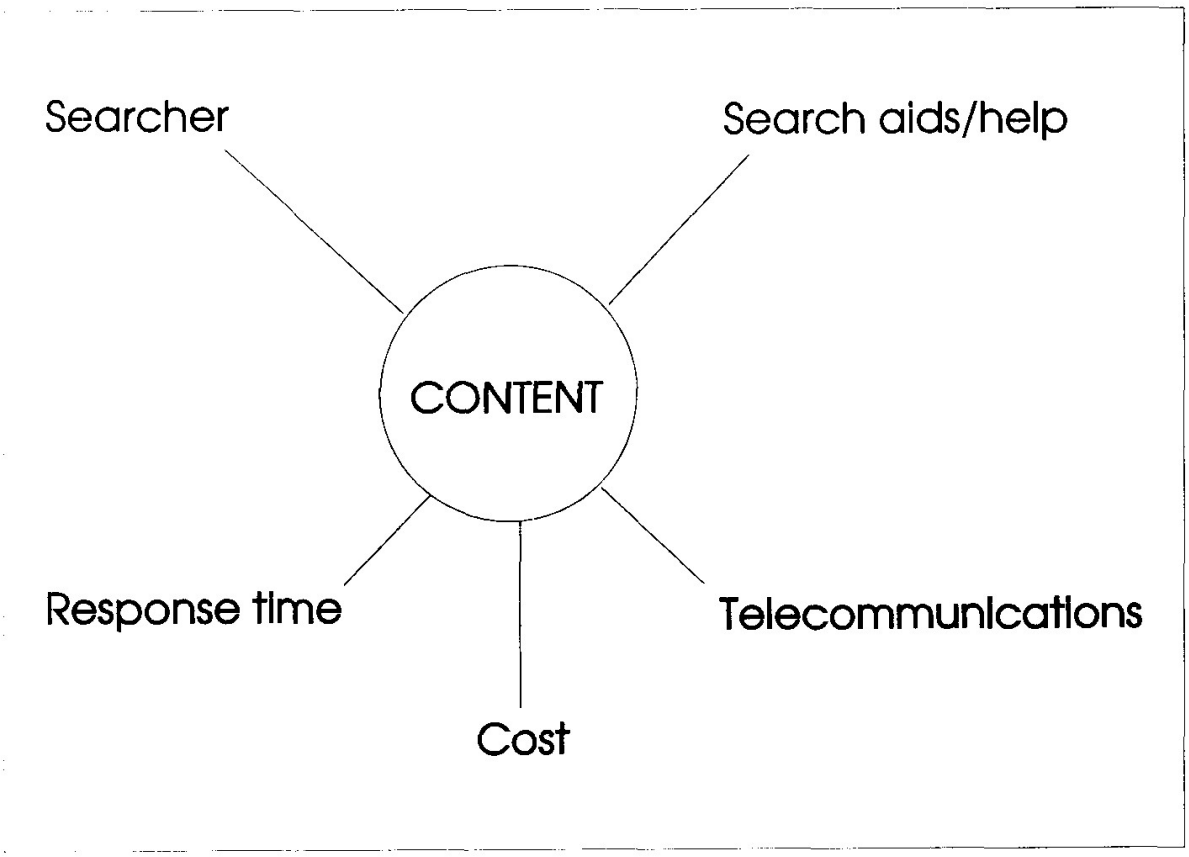

Figure 8.3 *Content is the core of quality*

> The [Information Access Company] databases are provided 'as is' without warranty of any kind. Neither IAC nor any of its data suppliers make any warranty whatsoever as to the accuracy or completeness of any of the databases or the results to be obtained from using the information contained therein and neither IAC nor any of its data suppliers shall be responsible for any claims attributable to errors, omissions or other inaccuracies in the information contained in any IAC database. The entire risk as to the results and performance of any IAC database is assumed by the user of that database ... .

Some producers warn users directly that their information may be of poor quality:

> INVESTEXT database users are advised that the electronic conversion and transmission of textual and numeric data may cause errors and/or omissions that are beyond the control of Thomson Financial Networks. You should also note that there is a delay between preparation of the research reports and their inclusion in the INVESTEXT database and the reports may therefore not be up-to-date.
>
> Teikoku Databank customer acknowledges that the type of information contained in the database will contain a degree of error. ...
>
> Elsevier Science Publishers does not warrant the scientific accuracy or

> completeness of the information contained in SEDBASE nor does it warrant the suitability of this information for any particular use or application. ... Because of rapid advances in the medical sciences, we recommend that independent verification of diagnoses and drug dosages should be made.

Let the poor buyer beware.

Disclaimers or not, database publishers such as Information Access Company, BIOSIS, Thomson, Elsevier and Teikoku are ultimately responsible (at least morally) for the quality, or lack thereof, of their content. Online hosts such as Dialog, Mead Data Central, Orbit and BRS are responsible too, only less directly so. They are responsible if their correction procedures are so awkward or expensive that they make it prohibitive for database producers to correct known errors. They are responsible if their loading programs do nothing to smooth out inconsistency of formats among and between databases.

Finally, searchers are responsible as well, if they continue to pay for junk data, or let errors go unreported, or fail to spread the word about content quality (both good and bad) of specific databases.

Content quality

What affects content quality? It is more than just typographical errors; as the Finnish Society for Information Services found, these may be relatively rare.[14] Here, the details from the individual quality checklists are helpful.

Analyzing the various checklists reveals that content commonly includes five things: (1) scope (also called coverage), (2) structure, (3) accessibility, (4) accuracy and (5) consistency. These five factors will be discussed in turn.

Scope

In a purely comparative list, *scope* can include many aspects, such as size, geographic coverage, language coverage, etc. In terms of quality, however, the SCOUG report summarized it best: no matter what a database has as its scope, quality in this regard means publishing a clear editorial policy and following that policy.[15]

A bibliographic database can restrict its coverage by only indexing journals in specified languages; a full-text database can be highly selective in the choice of articles it includes; a business directory database can set narrow criteria for which types of companies to include.

Narrow or restrictive scope does not mean poor quality, just as small size does not. Poor quality comes when the database publisher does not admit to those restrictions or does not clearly state the limitations. From the user's perspective, poor quality comes from not being told, or being told incorrectly.

For example, the BIOSIS database claims 'comprehensive worldwide coverage of

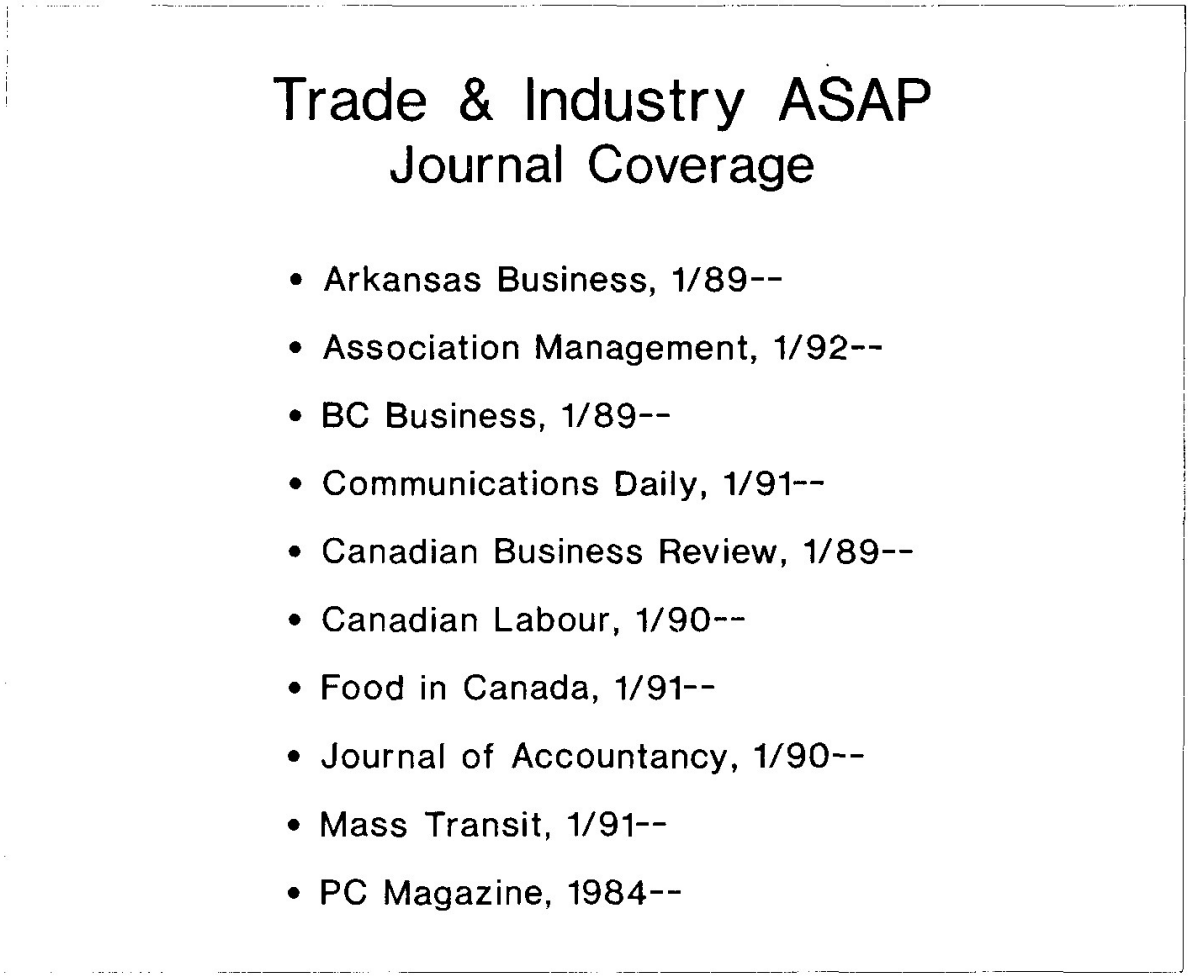

Figure 8.4 *Trade and Industry ASAP: journal coverage*

research in the life sciences' and *Compendex* claims 'coverage of the world's significant engineering and technological literature.' Yet a study revealed that scientific and technical journals published in Japan are only included in these two databases if the Japanese journal was written in English or another Western language or had English language summaries.[16]

Trade & Industry ASAP claims coverage of over 200 journals and a coverage timespan of 1983 to the present, but only 100 titles are covered for this entire period. As new titles are added, they are not added retrospectively, so some of the 200 titles begin coverage only in 1992 or 1993. Selected titles with their starting dates show this range (Figure 8.4).

National Newspaper Index's policy states that it 'provides cover-to-cover indexing of America's most important national newspapers,' yet it leaves out advertisements and some syndicated columns, not to mention the exclusions it does acknowledge, such as weather charts, stock market tables, crossword puzzles and horoscopes.

Very few, if any, bibliographic databases really do cover-to-cover indexing, although many claim they do. H.W. Wilson Company's advertising brochures claim cover-to-cover indexing for *Business Abstracts, Readers' Guide Abstracts* and other

indexes, but their system documentation admits that their policy is to select articles only. They do not index or abstract advertisements, most letters to the editor, calendar announcements or short items.

Structure

Structure, the second component of content, can be viewed in a similar way. Structure includes what, and how many, fields are in a database. Whether and how these fields can be searched is dependent on the loading and search software, but they must exist in the first place in order to be searchable. Lack of certain key fields has a strong negative effect on the quality of a database.

For instance, *Library and Information Science Abstracts* (LISA) has no article/document type indicator, so a search cannot be restricted by type of literature. *Dissertation Abstracts* indicates only the university that awarded a degree, not the department from which it came. Many databases do not include ISSN numbers or CODENs for the journals they index. Only some business databases include value-added fields such as SIC codes, stock market ticker symbols, etc.

The bottom line of structural quality, like that of scope, relates to honesty. In this case, honesty refers to what per cent of the records contain the promised fields. Jacsó's chapter in this book provides details on how to test field completeness.

If a database promises that one can search by language, but fewer than 5 per cent of the records have a language code, as Jacsó found for *Books in Print*, it is a misleading claim, to say the least.[17] If Library of Congress Card Number is advertised as a value-added feature (as *Books in Print* does), but this field appears in only 31 per cent of the records, using LCCN as a search element limits that search immediately to less than a third of the file.

Jacsó also found that 20 per cent of the records in *Disclosure* lack number-of-employees and net-sales-per-employee, and 15 per cent lack gross profits and net income.[18] Maybe that information could not be obtained from the companies, but nearly 15 per cent of the records are missing Standard Industrial Classification codes, something that could be added by the *Disclosure* staff.

ABI/INFORM has four fields that represent journal titles. Lest a searcher think that these fields can be used interchangeably to restrict a search to a specified journal, he or she should know that, while Journal Name and Journal Code appear in nearly 100 per cent of the records, and ISSN in 97 per cent, CODEN appears in only 62 per cent.

If certain fields appear only in some records in a database, the database producer has an obligation to report this. It should be clearly stated in all the documentation that describes field structure. No user should have to guess what per cent of the records can be expected to include a particular field. Failure to document such omissions is one of the worst quality transgressions of all – users miss items without knowing it because the database producer withheld information.

Accessibility

Like structure, *accessibility* varies according to the loading procedures and software that make a database searchable. But again, the core of quality – content – must be of high quality first. Where structural quality relates to the consistent availability of standard fields, accessibility extends to what potential additional access points are present in the database records.

MESH TREE STRUCTURE

HORMONES, SYNTHETIC	D6.597
ANDROGENS, SYNTHETIC	D6.597.100
ANABOLIC STEROIDS	D6.597.100.194
5-ANDROSTENE-3,17-DIOL	D6.597.100.194.50
ETHYLESTRENOL	D6.597.100.194.239
FLUOXYMESTERONE	D6.597.100.194.289
MESTEROLONE	D6.597.100.194.400
METHANDRIOL	D6.597.100.194.429
METHANDROSTENOLONE	D6.597.100.194.464
METHENOLONE	D6.597.100.194.500
NANDROLONE	D6.597.100.194.590
NORETHANDROLONE	D6.597.100.194.637
OXANDROLONE	D6.597.100.194.686
OXYMETHOLONE	D6.597.100.194.723
STANOZOLOL	D6.597.100.194.893
TRENBOLONE	D6.597.100.194.930

Figure 8.5 *MESH tree structure*

'Value-added' is the key to accessibility quality. In a textual database, either bibliographic or full-text, the presence of controlled vocabulary descriptors from a hierarchical thesaurus can improve both recall and precision.

For example, in *Medical Subject Headings* (MESH), the term Anabolic Steroids is a descriptor. Using this term will make a search more precise (it ensures that the searcher does not retrieve all the other types of steroids), and will increase recall because it will also pick up articles that refer to anabolic steroids by their chemical constituents only. In the MESH hierarchy, all of the specific anabolic steroids are displayed under the broad term 'anabolic steroids', so they can all be searched in a comprehensive search. They can be searched by name or by classification code (Figure 8.5).

Without descriptors or descriptor codes assigned from a hierarchical thesaurus, such sophisticated search features as Boolean OR, thesaurus viewing, explode (to retrieve automatically all of the narrower terms under a broad term) and truncation are unworkable, or far less useful.

As another example of how controlled vocabulary helps accessibility, hard-to-define concepts can be searched more easily with good controlled vocabulary. Consider a search on how things have changed for the people of Guam in the last decade. The user does not want everything on Guam, just information on how lifestyle, customs, income level, attitudes and outlook have changed. Those are difficult concepts to search free-text, but several databases, including *Sociological Abstracts* (Figure 8.6), contain the descriptor 'QUALITY OF LIFE', which sums up all of these ideas in one heading.

Luckily, database designers are rediscovering the value of thesauri as knowledge bases, after speculation by some that controlled vocabulary would become obsolete with full-text. Thesauri and controlled vocabulary indexing are not obsolete. Good indexing provides additional accessibility, improved results and additional database quality.

Other value-added fields can also improve accessibility. In bibliographic databases, abstracts provide increased retrieval points and help relevance judging. In full-text, abstracts bring all major concepts into focus in a single paragraph. Quality abstracts adhere to national or international standards, are as informative as possible, well written, and of an appropriate length (generally between 100 and 500 words).[19]

In directory databases, codes such as SIC numbers, Dun & Bradstreet (DUNS) numbers, Chemical Abstract Registry Numbers, etc. all aid the user if they are applied consistently. Authority control of company names, states, countries, etc. can greatly improve accessibility in directory databases as well.

If timeliness and comprehensive coverage of large amounts of information are essential to a product's value, a database producer can add even more value. After many years of development, the Institute for Scientific Information (ISI) began adding extra retrieval points to its records in the *Science Citation Index* and *Social Science Citation Index* databases.

These are both huge databases created almost completely by machine. ISI felt it could afford neither the time nor the cost to add human-indexed descriptors or

Quality of Life

SN Subjective evaluation of an individual's or group's way of life, lifestyle, or living conditions, usually using an explicit inventory of factors.

BT Quality
NT Quality of Working Life
RT Affluence
Deprivation
Disadvantaged
Everyday Life
Family Life
Humanization
Life
Life Satisfaction
Living Conditions
Social Conditions
Social Status
Socioeconomic Status
Standard of Living
Well Being

Figure 8.6 *Quality of life*

abstracts. Their Keywords+field does so automatically, by assigning keywords to records based on computer examination of the titles of articles listed in the footnotes of the article being indexed. Abstracts are added if the original article contains an author-written abstract.

Accuracy

The two final categories of top priority quality run throughout the other three categories, but also contain some unique properties of their own. *Accuracy* and *Consistency* are implied in high-quality editorial selection policies, field structures and indexing. The application of scope, structure and accessibility must, of course,

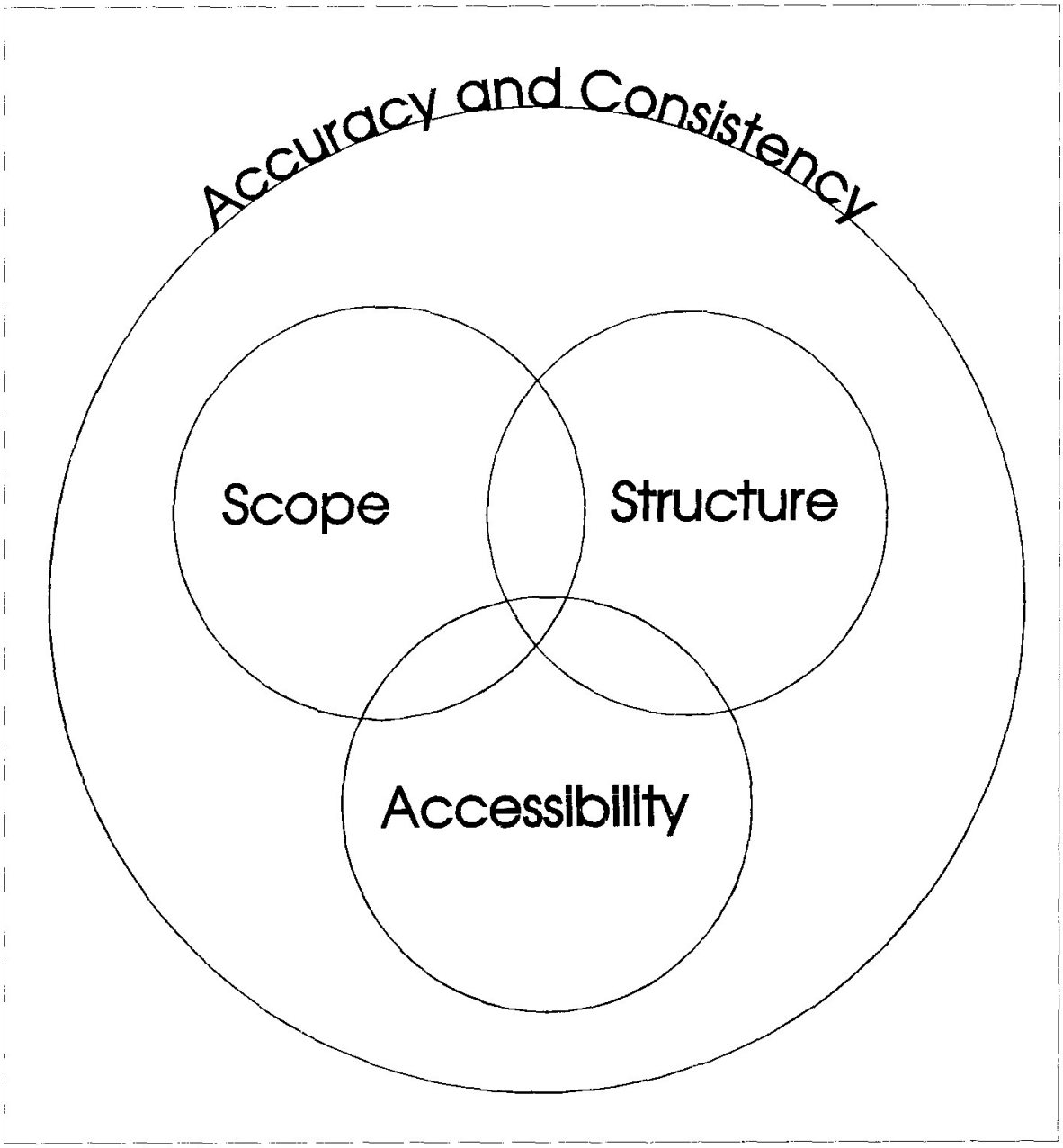

Figure 8.7 *Five aspects of content quality*

be both accurate and consistent to be of any use. Rather than a five-point list, the five aspects might be pictured as a Venn diagram (Figure 8.7), with accuracy and consistency included in all other content aspects. But accuracy and consistency are worth considering on their own. Accuracy is where typographical errors come to bear in the evaluation of content quality.

Accuracy is more important in some fields than in others. Factual and content-loaded fields should be double-checked for accuracy by database producers. Numeric fields, address fields, phone numbers and source fields, in particular, should be validated and verified, because the potential for problems in retrieval and use are so high if they contain errors.

For instance, Jacsó found that *Ulrich's International Serials Directory* listed the price of both *Biological Abstracts* and *Mathematical Reviews* as only $4 per year! Quite a bargain, but highly suspect.[20]

Typographical errors in text fields such as titles, abstracts, or full-text are unsightly, but do they impact search results adversely? Database producers claim that it is too expensive and time-consuming to spell-check and proofread to the degree necessary to get better accuracy. How devastating are typos in these fields?

Part of the answer to that question has to do with statistical probabilities. The more information there is in a database record, the higher the probability that the error will not be devastating. Thus, in a full-text database, the odds that one or two

Vacuums in INSPEC

```
17915  VACUUM
    2  VACUUN
    1  VACUU
    1  VACUUJM
    2  VACCUMM
    1  VACCUMS
  106  VACCUUM
    1  VACCUUMS
    1  VACUUMJ
    5  VACUUMP
  737  VACUUMS

S VACCUUM? AND VACUUM? = 31
S VACCUUM? NOT VACUUM? = 76
```

Figure 8.8 *Vacuums in INSPEC*

typos will adversely affect the search results is small, since the words are likely to be repeated elsewhere in the record.

This redundancy factor, or its lack, can be easily illustrated. Figure 8.8 shows a list of the variety of spellings of 'vacuum' in INSPEC. Out of the 107 misspellings *Vaccuum* or *Vaccuums*, 31 of the records also contain the word spelled correctly. This means that 76 of the records do not have any redundancy and would therefore be lost in a search if the searcher did not first check the index and choose all incorrect spellings along with the correct one.

A 1977 study by Charles Bourne pointed out that forcing a searcher to do such checking increases the cost of the search and necessitates more time to do a comprehensive search.[21] He noted that, even then, only errors that begin with the same word stem are likely to be retrieved, either from an index display or from truncation. Words with the first letters wrong are likely to go unnoticed by users.

Generally, redundant errors will adversely affect retrieval only if the user is searching with proximity operators and the word pattern is not repeated anywhere else in the document. Publishers are often willing to take this risk with searchers' results.

Imagine walking into a bookstore and seeing a sign: 'Books with errors and typos $20 – Proofread books $50'. Consumers would not stand for it, but database producers get away with such practices regularly. Many database producers claim that users cannot afford to pay what it would take to have cleaner databases, so they must be satisfied with the products as they are.

Non-redundant errors are especially problematic. These occur more often in short fields than in lengthy texts or abstracts. The 'parent country' field in *D&B – International Dun's Market Identifiers*, shown below (Figure 8.9), contains just one value. Someone looking for all records about England will be bound to miss many directory listings. An even more serious accuracy problem, especially with directory and numeric databases, is the reliability of the source information. In a company database record, if the name of the president, the phone number, or the gross sales are inaccurate because of poor data-gathering techniques, the record, and by implication the entire database, is worthless.

The Supreme Court case of Dun & Bradstreet (D&B) *vs.* Greenmoss Builders has been widely reported. A student hired by D&B to go through and interpret legal documents mistakenly reported that Greenmoss had filed for bankruptcy. Greenmoss sued D&B and won.[22]

Another well-known example is when the *Dun & Bradstreet European Market Identifiers* database was found to have used inaccurate currency conversion formulas. They alternately reported three Yugoslavian and three Greek companies as being the largest food companies in the world.[23]

Dun & Bradstreet's reaction to these problems was, in part, to revise their disclaimer. On Dialog, the *D&B Dun's Financial Records Plus* disclaimer statement reads in part: 'You acknowledge that D&B does not warrant or guarantee the timeliness, currentness, accuracy, completeness, merchantability or fitness for a particular purpose of the information. You also acknowledge that every business

decision involves the assumption of a risk and that D&B, in furnishing the information to you, cannot and will not underwrite that risk in any manner whatsoever. You therefore agree that D&B will not be liable for any loss, damage or injury caused in whole or in part by D&B's negligence in procuring, compiling, collecting, interpreting, reporting, communicating or delivering the information... ."[24]

If the wrong information is reported, no amount of data input checks and double checks will ensure accuracy. Unfortunately, it is difficult for consumers to verify this aspect of accuracy, unless the correct information or source of the information is already known to them.

Because of this uncertainty, searchers must always question the factual information they retrieve. If a directory database includes large numbers of records with information listed as NA (not available), one must question the database's techniques for data verification. If something looks suspicious an experienced searcher will try to verify it in another source.

A corporate librarian in an oil trading firm reported that she always spot checks the worldwide oil price data they download every day. After doing this for 15 years, she recognizes when a price 'just doesn't sound right.' For those, she calls other sources, then adds a note to the price on their in-house system, alerting the traders

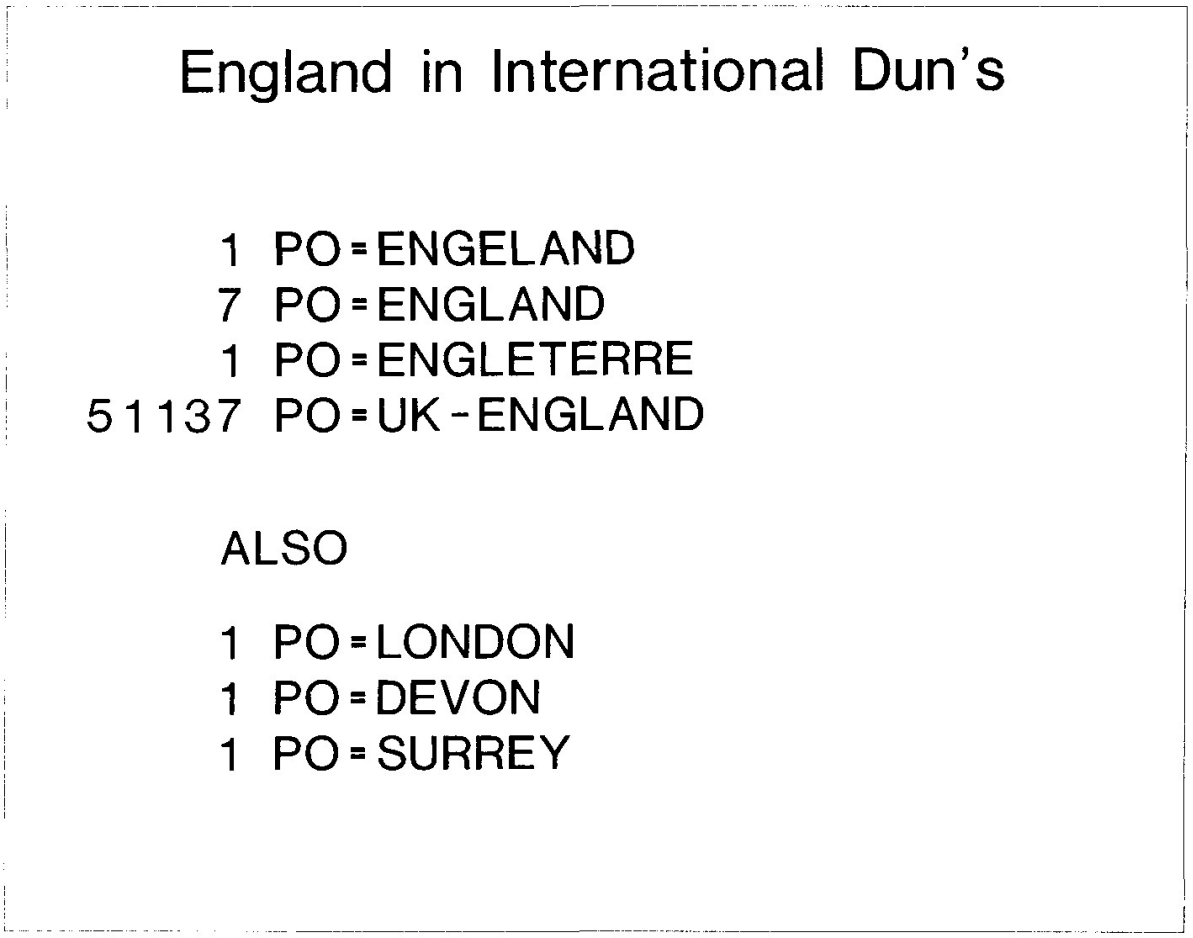

Figure 8.9 *England in International Dun's*

to the verified information. Since no other companies bother to do this, hers has an advantage on the trading floor, because only her traders know the real price. Unfortunately, there is no better way to verify this type of data.

Consistency

Finally, probably the single most important measure of quality is consistency. A user can live with a database of limited scope if the producer publicizes the limitations and follows the statement of scope consistently. He or she can live with a database that has few fields if all are consistently present in the records. There may even be some merit in consistently making the same error.

Consistency allows predictability. Consistency helps users realize what they will and will not get in a database search. It makes search aids, such as manuals and help messages, meaningful.

Consistency cuts across all parts of a database. For example, a researcher who is running updates to a search on a regular basis needs to be able to rely on consistency in such matters as when a database is updated, the approximate number of records added at each update, the terminology used to index the subject, and the fields available for searching.

INVESTEXT is supposed to be updated semi-weekly on Dialog. Sometimes it is; other times there is as long as a three-week lag between updates.[25] Articles in ERIC about how schools can help prevent child abuse consistently use the descriptor *Child Abuse*, but variously use *School Role*, *School Responsibility*, *Teacher Role*, *Teacher Responsibility*, *Prevention*, and others, to index the other concept.

All of these are consistency problems. All greatly impact the quality of search results.

Advice to database producers

Database producers can improve content, the most important aspect of quality, to make quality products. Database producers have an obligation to provide an acceptable level of content quality while maintaining currency. It may not be easy, but at some point users are not going to continue to accept excuses.

Here are some fundamental techniques for improving database quality at the producer end. Some customers may be surprised that these are not being widely employed already:

1. Use spell-checkers with verification to correct scanning and data input errors
2. Use data validation on structured fields
3. Spend time on thesaurus development, maintenance and adding value
4. Hire qualified personnel to gather, index, input, check, etc.
5. Put pressure on online services to improve their error correction procedures
6. Add easy mechanisms for customers to report errors (such as Mintz's FIXIT

command, first adopted by NewsNet)[26]
7. Develop, print, and publicize editorial policies *and follow* them
8. Believe that users need and want content quality, and understand that poor practices cannot go on forever
9. If you can not manage a reasonable level of content quality, get out of the business (searchers should help by refusing to buy poor quality products).

Certainly, there are trade-offs. Time lags may be a problem already, without spending even more time and care in preparation. But as more products become available in electronic form, and as a wider base of consumers uses them, poor quality will not continue to sell. Better software for scanning, error detection, validation and automatic indexing can help with many of these quality issues, although not with all.

Finding out about quality

All searchers share the responsibility for testing and reporting on the quality of database content. If searchers do not make it their job to test quality, they will end up having to issue the same sort of disclaimers as database producers – but theirs will read 'let the client beware'.

Testing is not as difficult as many searchers imagine; not everyone need become a full-time quality control officer. There is a continuum of testing, from simple to complex, from quick to more time-consuming, from inexpensive to potentially quite expensive. Many more searchers will find a role for themselves on the quick or simple end than on the complex end. Each searcher can decide where on the continuum to draw the line of direct involvement (Figure 8.10).

Jacsó's chapter provides methods and details of many quality tests along all points of the continuum. The simplest technique, however, is one that every searcher can perform routinely for every database he or she searches. This technique is simply to see if someone else has already tested the quality of the database and written a review. Every searcher should be in the habit of regularly reading database reviews. Of course, one must be careful of who writes the reviews – an employee of the database producer is hardly likely to be unbiased. Some journals label these product reviews in some way, such as 'First Look', but with others the reader has to check author affiliation to make sure the review is objective. There are many good sources of database reviews. A selected source list is included at the end of this chapter.

A consortium such as the Finnish Society for Information Services, SCOUG, or any other group of concerned users, is a good avenue to accomplish systematic database tests on a large scale. Large-scale testing is prohibitive for most individuals, due to the time and money required to test a database rigorously. Working together, however, searchers could create a regular and consistent database reviewing vehicle. Some groups have proposed a sort of international *Consumer Reports* of databases.

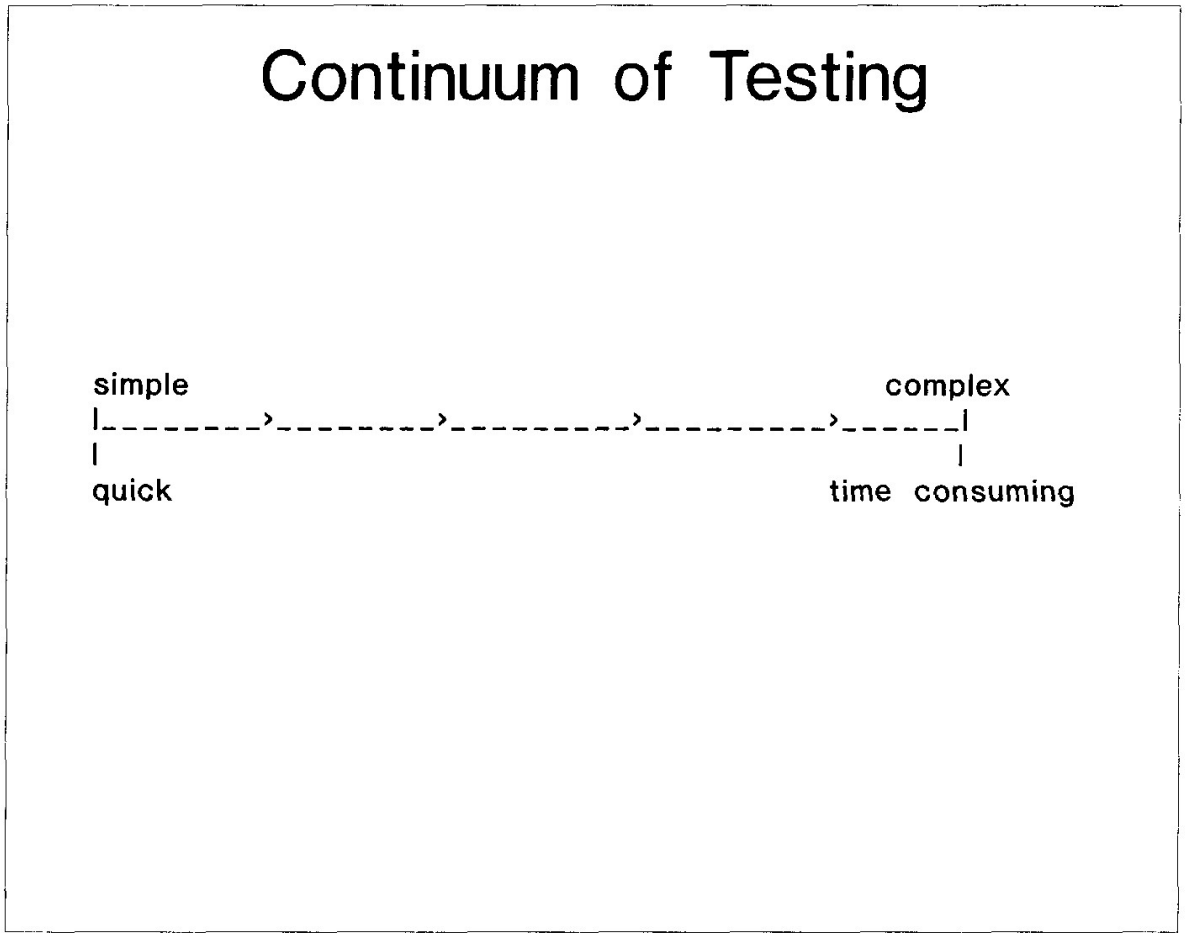

Figure 8.10 *Continuum of listing*

Internet is a natural distribution vehicle for such reports. Reports could be built incrementally, with many searchers from all over the world adding their input for each database. The development of a consistent checklist and standardized testing methods would assure some uniformity.

Conclusion

Content, made up of consistent and accurate scope, structure, and accessibility, is the top priority for quality. All quality is not created equal: without quality content, no amount of fancy programming, nifty interfaces, or clean telecommunications lines makes any difference. Database producers can create higher quality databases if they are convinced that customers care. Often, poor content quality is hidden, because it is never tested systematically. Jacsó's chapter provides several systematic methods for tests that measure database quality and that can help uncover some of the hidden dirty secrets of database content.

Selected sources for evaluating databases

Gale Directory of Databases. Detroit, MI: Gale Research, Inc. 2 volumes. Twice yearly. January 1993–. This merges and replaces three popular directories: *Computer-Readable Databases: A Directory and Data SourceBook*, Martha E. Williams, Founding Editor; *Directory of Online Databases and Directory of Portable Databases*, founded by Cuadra Associates. Volume One covers online databases, Volume Two covers CD-ROM, diskette, tapes, handheld, and other 'portable' formats.

Fulltext Sources Online. Ruth Orenstein, editor. Needham Heights, MA: BiblioData, twice yearly. A directory of periodicals (magazines, journals, newspapers, newsletters, and newswires) available online in full-text from major international vendors.

Jacsó, Peter. CD-ROM Software, *Dataware, and Hardware: Evaluation, Selection, and Installation*. Englewood, CO: Libraries Unlimited, 1992. Examples are for CD-ROM databases only, but the many useful techniques to test and evaluate databases work for any electronic distribution medium.

Database magazine (Online Inc., Wilton, CT). Probably the best source for comparative reviews and techniques for testing.

Online magazine is the companion to *Database* (they are published every other month by the same publisher). It has fewer database reviews, but many system reviews and discussions of broader issues of quality.

Online Review (Learned Information, Ltd., Oxford, England). Publishes reviews of databases and systems and research articles, as well as database news.

Information Today (Learned Information, Inc., Medford, NJ). Industry news, columns and some database reviews. Good for announcements of new databases and movements to improve quality.

Searcher: The Magazine for Database Professionals. Barbara Quint, editor (Learned Information Inc., Medford, NJ). Quint is an outspoken proponent of quality searchers and the guiding light of the Southern California Online Users Group.

RQ (American Library Association, Reference and Adult Services Division, Chicago, IL). Includes reviews of online and CD-ROM databases.

Notes

1. For a summary, see: Tenopir, Carol (1992), 'Evaluation criteria for online, CD-ROM', *Library Journal* **117** (March 1): 66,68; Tenopir, Carol, 'Database quality revisited', *Library Journal* **115** (October 1), 64–5, 67.
2. Basch, Reva (1990), 'SCOUG retreat addresses data quality issues', *Online* **14** (November), 81–2.

3. Juntunen, Ritva *et al.*, (1991), 'Quality requirements for databases – Project for evaluating Finnish databases', in *Online Information '91*, 15th International Online Information Proceedings, London, 10–12 December 1991 (Ed. David I. Raitt), Oxford, England: Learned Information Ltd. 351–9.
4. Jacsó, Peter (1992), *CD-ROM Software, Dataware and Hardware: Evaluation, Selection and Installation*, Englewood, Colorado: Libraries Unlimited.
5. Tenopir, 'Database quality revisited', 64.
6. Juntunen, 357–9.
7. Jacsó, *CD-ROM Software, Dataware and Hardware: Evaluation, Selection and Installation.*
8. Williams, Martha E. (1990), 'Highlights of the online database industry and the quality of information and data', *National Online Meeting Proceedings*, 1990, New York, May 1–3 (Ed. Martha E. Williams)/Medford, New Jersey: Learned Information, Inc.
9. Tenopir, 'Database quality revisited', 64.
10. Jacsó, *CD-ROM Software, Dataware and Hardware: Evaluation, Selection and Installation.*
11. Tenopir, 'Evaluation criteria for online, CD-ROM', 66.
12. Juntunen, 357–9.
13. Database Supplier Terms and Conditions (Palo Alto, California, Dialog Information Services, Inc., July 1991)
14. Juntunen, 354.
15. Tenopir, 'Database quality revisited', 64.
16. Ikushima, Keiko and Tenopir, Carol (1988), 'Availability of Japanese scientific and technical periodicals in major English language databases', in *National Online Meeting Proceedings*, 1988, New York, May 10–12 (Eds Martha E. Williams and Thomas H. Hogan), Medford, New Jersey, Learned Information Inc., 115–22.
17. Jacsó, Peter, Workshop held at National Online Meeting, New York, May 1992.
18. Ibid.
19. Tenopir, Carol and Jacsó, Peter, 'Quality of Abstracts', *Online* (May 1993), 44–55.
20. Jacsó, *CD-ROM Software, Dataware and Hardware: Evaluation, Selection and Installation*, 136.
21. Bourne, Charles P. (1977), 'Frequency and Impact of Spelling Errors in Bibliographic Databases', *Information Processing and Management* **13**, 10.
22. Mintz, Anne (1985), 'Information Practice and Malpractice', *Library Journal* **110**, 15 September, 43.
23. Pagell, Ruth (1990), 'Sorry, Wrong Number: Exchange Rates and Sales Figures', *Online* **14** (November), 20–1; Ruth Pagell (1991), 'It's Greek to Me! Exchange Rate Translations and Company Comparisons', *Database* **14** (February), 21–7.
24. Database Supplier Terms and Conditions, op. cit.
25. Tenopir, Carol and Hover, Katie (1993), 'When is the same database not the

same?' *Online* (July), 20–7.
26. Mintz, Anne (1990), 'Quality Control and the Zen of Database Production', *Online* **14** (November), 17.